



H-1141

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

K. SERIZAWA et al.

Serial No. 10/825,158

Group Art Unit: 2188

Filed: April 16, 2004

Examiner: M. PADMANABHAN

For: METHOD FOR ALLOCATING STORAGE
AREA TO VIRTUAL VOLUME

TRANSMITTAL OF CERTIFIED PRIORITY DOCUMENT

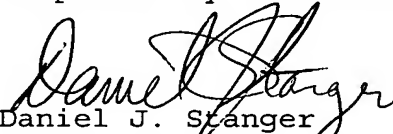
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

February 9, 2005

Sir:

Submitted herewith is a certified priority document
(JP 2004-063392) of a corresponding Japanese patent
application for the purpose of claiming foreign priority under
35 U.S.C. § 119. An indication that this document has been
safely received would be appreciated.

Respectfully submitted,


Daniel J. Stanger
Registration No. 32,846
Attorney for Applicants

MATTINGLY, STANGER, MALUR & BRUNDIDGE, P.C.
1800 Diagonal Rd., Suite 370
Alexandria, Virginia 22314
(703) 684-1120
Date: February 9, 2005

日 本 国 特 許 庁

JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 4 年 3 月 8 日
Date of Application:

出 願 番 号 特 願 2 0 0 4 - 0 6 3 3 9 2
Application Number:

ST. 10/C] : [J P 2 0 0 4 - 0 6 3 3 9 2]

出 願 人 株 式 会 社 日 立 製 作 所
Applicant(s):

2 0 0 4 年 1 0 月 2 1 日

特許庁長官
Commissioner,
Japan Patent Office

小 川

洋

CERTIFIED COPY OF
PRIORITY DOCUMENT

BEST AVAILABLE COPY

出証番号 出証特 2 0 0 4 - 3 0 9 5 0 8 8

【書類名】 特許願
【整理番号】 K04004681A
【あて先】 特許庁長官殿
【国際特許分類】 G06F 12/00
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
 システム開発研究所内
 【氏名】 芹沢 一
【発明者】
 【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所
 ソフトウェア事業部内
 【氏名】 森田 眞司
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
 システム開発研究所内
 【氏名】 岩見 直子
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社 日立製作所
【代理人】
 【識別番号】 100075096
 【弁理士】
 【氏名又は名称】 作田 康夫
【選任した代理人】
 【識別番号】 100100310
 【弁理士】
 【氏名又は名称】 井上 学
【先の出願に基づく優先権主張】
 【出願番号】 特願2003-150082
 【出願日】 平成15年 5月28日
【手数料の表示】
 【予納台帳番号】 013088
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1
 【包括委任状番号】 9902691
 【包括委任状番号】 0315335

【書類名】 特許請求の範囲**【請求項 1】**

各々制御部とディスク装置とを有する複数のストレージシステムと、
前記複数のストレージシステムと接続される仮想化装置とを有するシステムであって、
前記仮想化装置は、計算機と接続される入力ポートと、ストレージシステムと接続される出力ポートと、入力ポートと出力ポート間でデータを転送する転送部と、制御部と、計算機からのアクセス要求の対象となる仮想ボリュームと該仮想ボリュームに割り当てられている前記複数のストレージシステム内の記憶領域との対応関係を記録したアクセス変換テーブルとを有しており、

前記制御部は、計算機から仮想ボリュームに対する書き込み要求を受信した際に、該書き込み要求に応じて、前記複数のストレージシステムが有する記憶領域を前記仮想ボリュームに割り当てると共に前記アクセス変換テーブルを更新し、

更新後の前記アクセス変換テーブルに基づいて、計算機から受信した前記書き込み要求を、前記仮想ボリュームに新たに割り当てられた記憶領域を有するストレージシステム宛ての書き込み要求に変換し、

前記転送部は、前記仮想ボリュームに新たに割り当てられた記憶領域を有するストレージシステムに対して、変換後の書き込み要求を送信するよう制御することを特徴とするシステム。

【請求項 2】

計算機に接続される仮想化装置と、

前記仮想化装置に接続される複数のストレージ装置とを有し、

前記仮想化装置は、前記計算機からの要求に基づいて、前記計算機に所定の大きさの仮想ボリュームが割り当てられた旨の通知を発行し、

前記計算機から前記仮想ボリュームに対するアクセス要求を受信した際に、前記複数のストレージ装置内に存在する記憶領域を前記仮想ボリュームに割り当てて、前記計算機から受信したアクセス要求を、前記仮想ボリュームに割り当てた記憶領域を有するストレージ装置宛てのアクセス要求に変換して、該ストレージ装置に送信することを特徴とするシステム。

【請求項 3】

請求項 2 記載のシステムであって、

前記所定の大きさの仮想ボリュームが割り当てられた旨の通知には、仮想ボリューム識別子と前記仮想ボリュームのサイズとが含まれており、

前記計算機から前記仮想ボリューム識別子を有するアクセス要求を受信した際に、前記仮想化装置は、該仮想ボリューム識別子によって特定される仮想ボリュームに前記複数のストレージ装置内に存在する記憶領域を割り当てて、前記計算機から受信したアクセス要求を、該仮想ボリュームに割り当てられた記憶領域を有するストレージ装置宛てのアクセス要求に変換して、該ストレージ装置に送信することを特徴とするシステム。

【請求項 4】

請求項 3 記載のシステムであって、

前記仮想化装置は、仮想ボリュームと該仮想ボリュームに割り当てられている前記複数のストレージ装置内の記憶領域との対応関係を示すアクセス変換テーブルを有しており、

前記仮想ボリュームに前記複数のストレージ装置内に存在する記憶領域を割り当てた場合に、前記アクセス変換テーブルを更新することを特徴とするシステム。

【請求項 5】

請求項 4 記載のシステムであって、

前記仮想化装置は、更新後のアクセス変換テーブルを参照して前記計算機から受信したアクセス要求を変換し、変換後のアクセス要求をストレージ装置に送信することを特徴とするシステム。

【請求項 6】

請求項 4 記載のシステムであって、

前記仮想化装置は、前記アクセス変換テーブルを更新して、仮想ボリュームに割り当てられている記憶領域のうち、前記計算機に使用されなくなった記憶領域の割り当てを解放することを特徴とするシステム。

【請求項 7】

請求項 3 記載のシステムであって、

前記仮想化装置は、ファイルシステムの初期化に基づく管理データの書き込み時には、前記計算機からアクセス要求を受信した際より小さい容量の記憶領域を前記仮想ボリュームに割り当てて、割り当てられた記憶領域に管理データを書き込むことを特徴とするシステム。

【請求項 8】

請求項 6 記載のシステムであって、

前記仮想化装置は、ファイルシステムのログ情報を利用して、仮想ボリュームに割り当てられている記憶領域のうち、前記計算機に使用されなくなった記憶領域を特定することを特徴とするシステム。

【請求項 9】

請求項 3 記載のシステムであって、

前記仮想化装置は更に、仮想ボリュームに割り当て可能な前記複数のストレージ装置内に存在する空き記憶領域の容量が記録された空き容量管理テーブルを有しており、前記計算機に該空き容量管理テーブルに記録されている情報を出力することを特徴とするシステム。

【請求項 10】

請求項 9 記載のシステムであって、

前記仮想化装置は、仮想ボリュームに前記複数のストレージ装置内に存在する記憶領域を割り当てた場合に、前記空き容量管理テーブルを更新することを特徴とするシステム。

【請求項 11】

複数のストレージ装置と、該複数のストレージに格納されているデータにアクセスする計算機に接続される仮想化装置であって、

計算機に接続される第一のポートと、

ストレージ装置に接続される第二のポートと、

第一のポートと第二のポートとの間のデータの転送を実行する転送部と、

制御部とを有しており、

前記第一のポートは計算機から仮想ボリュームに対する書き込み要求を受信し、

前記制御部は、受信した書き込み要求に基づいて、前記仮想ボリュームに前記複数のストレージ装置に存在する記憶領域を割り当てて、受信した書き込み要求を前記仮想ボリュームに割り当てられた記憶領域を有するストレージ装置宛ての書き込み要求に変換し、

前記転送部は、変換後の書き込み要求を、前記仮想ボリュームに割り当てられた記憶領域を有するストレージ装置と接続している第二のポートを介して該ストレージ装置に送信するよう制御することを特徴とする仮想化装置。

【請求項 12】

請求項 11 記載の仮想化装置であって、

仮想化装置は、計算機からの要求に従って、前記仮想ボリュームを識別するための仮想ボリューム識別情報と、前記仮想ボリュームのサイズとを、前記計算機に通知し、

前記計算機から受信する書き込み要求には、前記仮想ボリューム識別情報が含まれることを特徴とする仮想化装置。

【請求項 13】

請求項 12 記載の仮想化装置であって、

前記仮想化装置は更に、仮想ボリュームと該仮想ボリュームに割り当てられている前記複数のストレージ装置内の記憶領域との対応関係を示すアクセス変換テーブルを有しており、

前記制御部は、前記仮想ボリュームに前記複数のストレージ装置内に存在する記憶領域

を割り当てた場合に、前記アクセス変換テーブルを更新することを特徴とする仮想化装置。

【請求項 14】

請求項 13 記載の仮想化装置であって、
前記制御部は、更新後の前記アクセス変換テーブルを参照して前記計算機から受信した書き込み要求を変換することを特徴とする仮想化装置。

【請求項 15】

請求項 13 記載の仮想化装置であって、
前記制御部は、前記アクセス変換テーブルを更新して、仮想ボリュームに割り当てられている記憶領域のうち、前記計算機に使用されなくなった記憶領域の割り当てを解放することを特徴とする仮想化装置。

【請求項 16】

請求項 12 記載の仮想化装置であって、
前記制御部は、ファイルシステムの初期化に基づく管理データの書き込み時には、前記計算機から書き込み要求を受信した際より小さい容量の記憶領域を前記仮想ボリュームに割り当てて、割り当てられた記憶領域に管理データを書き込むことを特徴とする仮想化装置。

【請求項 17】

請求項 15 記載の仮想化装置であって、
前記制御部は、ファイルシステムのログ情報を利用して、仮想ボリュームに割り当てられている記憶領域のうち、前記計算機に使用されなくなった記憶領域を特定することを特徴とする仮想化装置。

【請求項 18】

請求項 12 記載の仮想化装置であって、
前記仮想化装置は更に、仮想ボリュームに割り当て可能な前記複数のストレージ装置内に存在する空き記憶領域の容量が記録された空き容量管理テーブルを有しており、前記計算機からの要求に応じて、前記空き容量管理テーブルに記録されている情報を出力することを特徴とする仮想化装置。

【請求項 19】

請求項 18 記載の仮想化装置であって、
前記制御部は、仮想ボリュームに前記複数のストレージ装置内に存在する記憶領域を割り当てた場合に、前記空き容量管理テーブルを更新することを特徴とする仮想化装置。

【書類名】明細書

【発明の名称】記憶領域割当方法、システム及び仮想化装置

【技術分野】

【0001】

本発明は、記憶装置システムが有する記憶領域の計算機への割り当て技術に関する。

【背景技術】

【0002】

複数の記憶装置システムが有する複数の記憶領域をまとめて、一つの記憶領域として計算機にアクセスさせるスイッチが特許文献1に開示されている。特許文献1によれば、スイッチは複数の記憶装置システムとホストとに接続されている。スイッチはホストから、複数の記憶装置システムが有する複数の記憶領域（論理ユニット）から構成される統合論理ユニットに対するアクセス要求を受信する。そしてスイッチは、受信したアクセス要求を、アクセス対象のデータが格納されている論理ユニットに対するアクセス要求に変換して、この論理ユニットを有する記憶装置システムに変換後のアクセス要求を送信する。

【0003】

一方、計算機に割り当てられている論理的な記憶領域が拡大された場合に、記憶領域の拡大を計算機側で認識するためのソフトウェア技術は、非特許文献1に開示されている。

【0004】

【特許文献1】特開2000-242434号公報

【非特許文献1】“エルヴィエム ハウツウ9.9 (LVM HOWTO9.9)”、[online]、リナックスオンライン(Linux online)、[平成16年2月10日検索]、インターネット<URL:<http://www.linux.org/docs/ldp/howto/LVM-HOWTO/x592.html>>

【発明の開示】

【発明が解決しようとする課題】

【0005】

計算機に割り当てられている記憶領域のサイズを拡大する場合、非特許文献1に示すように、計算機が記憶領域のサイズの拡大を認識するためのソフトウェアが必要となる。従って、記憶装置システムに計算機を接続して、計算機が記憶装置システム内の記憶領域にアクセスしてこの記憶領域に格納されたデータを使用するシステムにおいて、計算機に割り当てられた（即ち、計算機からアクセス可能な）記憶領域のサイズを変更する場合には、計算機にこのソフトウェアをインストールしておく必要がある。

【0006】

しかし、記憶装置システムに複数の計算機が接続されている場合には、複数の計算機全てにこのソフトウェアを追加する必要がある、また複数の計算機の種類が各々異なる場合には計算機の種類に応じたソフトウェアをインストールする必要がある。さらに、計算機で使用されるオペレーティングシステムがバージョンアップした場合に、それに追従して、対応するソフトウェアを更新する必要がある。従って、計算機に記憶領域のサイズの拡大を認識させるためにソフトウェアをインストールしておくのは、管理負荷が大きく、コストも高くなる。

【0007】

ところで、計算機に割り当てられている記憶領域のサイズが拡大されるのは、計算機が扱うデータ量が増大し、当初割り当てられていた記憶領域では記憶容量が足りなくなってしまうからである。従って最初から大きなサイズの記憶領域を計算機に割り当てておけば、記憶容量の不足は起こらず、途中で計算機に割り当てられている記憶領域のサイズを拡大する必要もないので、計算機に上述の様なソフトウェアをインストールしておく必要もない。しかしそうすると、計算機がそれほど大量のデータを使わない場合には、計算機に割り当てられた記憶領域に未使用領域が発生して記憶領域に無駄が生じる。

【0008】

そこで、計算機に対して効率的に記憶領域を割り当てることが可能なシステムを開示する。

【課題を解決するための手段】**【0009】**

システムは、計算機に接続される仮想化装置と、仮想化装置に接続される複数の記憶装置とを有する。仮想化装置は、計算機からの要求に応じて、計算機に所定の大きさの仮想ボリュームを割り当てた旨の通知を発行する。計算機から仮想ボリュームに対するアクセス要求を受信した際に仮想化装置は、複数の記憶装置内に存在する記憶領域を仮想ボリュームに割り当て、計算機から受信したアクセス要求を、仮想ボリュームに割り当てた記憶領域を有する記憶装置宛てのアクセス要求に変換して、この記憶装置に変換後のアクセス要求を送信する。

【発明の効果】**【0010】**

本発明によれば、効率的に計算機に記憶領域を割り当てることができる。

【発明を実施するための最良の形態】**【0011】**

以下、本発明の実施形態を、図面を用いて説明する。尚、本発明は以下に説明する実施形態に限定されるものではない。

【実施例1】**【0012】**

図1は、本発明を適用したシステムの、第一の実施形態の構成例を示す図である。

【0013】

システムは、少なくとも1台のホストプロセッサ12、少なくとも1台のストレージ装置13、仮想化スイッチ11及び管理コンソール14を有する。

【0014】

ホストプロセッサ12は、ストレージ装置13に格納されたデータを使用する計算機である。ホストプロセッサ12は、仮想化スイッチ11が提供する記憶領域を仮想化スイッチ11に接続されていない他の計算機に提供する機能を有するファイルサーバでも良い。

【0015】

ストレージ装置13は記憶装置又は記憶装置システムである。ここで、記憶装置とは、ハードディスクドライブやDVDドライブ等の単体の記憶装置を指し、記憶装置システムとは、制御部及び制御部に接続された一又は複数のハードディスクドライブ等のディスク装置を有し、制御部が各ディスク装置への入出力を制御するストレージサブシステムを指すものとする。尚記憶装置システムは、複数のハードディスクドライブがRAID構成を有しているディスクアレイシステムであっても良い。

【0016】

ストレージ装置13は、少なくとも1つの論理ユニット（以下「LU」と呼ぶ。）131を有する。LU131は、ストレージ装置13が有する物理的な記憶領域から構成されている論理的な記憶領域である。LU131は、ホストプロセッサ12などストレージ装置13に接続される装置には、論理的に独立した1つのストレージ装置として認識される。

【0017】

また、LU131は、複数の部分的な論理的記憶領域（以下「実領域」と呼ぶ。）132から構成される。実領域132の各々も、ストレージ装置13が有する物理的な記憶領域に対応する。実領域132のサイズは任意であり、実領域132は、連続したアドレスをもつ領域である。

【0018】

仮想化スイッチ11は、図示するように通信線やスイッチによって他の装置（ホストプロセッサ12、管理コンソール14、若しくはストレージ装置13）と接続されており、他の装置と通信することができる装置である。また、仮想化スイッチ11は、仮想化スイッチ11に接続されている一又は複数のストレージ装置13が有する一又は複数の記憶領域をまとめて一の記憶領域として管理する（一又は複数のストレージ装置が有する一又

は複数の記憶領域をまとめて、一の記憶領域として管理することを、「仮想化する」と称する) 仮想化装置である。そして仮想化スイッチ 11 は、仮想化された記憶領域に対するアクセス要求をホストプロセッサ 12 から受け付けると、この仮想化された記憶領域に対応するストレージ装置 13 内の記憶領域を特定し、受信したアクセス要求を特定された記憶領域に対するアクセス要求に変換して、変換後のアクセス要求をストレージ装置 13 に送信する (仮想化された記憶領域に対するアクセス要求を受領して、これに対応するストレージ装置 13 内の記憶領域へのアクセス要求をストレージ装置 13 に送信することにより、仮想化された記憶領域に対するアクセス要求を処理することを、「仮想化された記憶領域を提供する」と呼ぶ)。

【0019】

仮想化スイッチ 11 とホストプロセッサ 12 との間、および仮想化スイッチ 11 とストレージ装置 13 との間で使用される通信線やスイッチでは、ファイバチャネル等のプロトコルが使用される。ただし、使用される通信線やプロトコルはローカルエリアネットワーク等で使用される通信線やプロトコルでも良い。

【0020】

仮想化スイッチ 11 は、ホストプロセッサ 12 とストレージ装置 13 との間に接続され、ホストプロセッサ 12 が発行するコマンドをストレージ装置 13 側に転送する機能を有する。また、仮想化スイッチ 11 がホストプロセッサ 12 に対して提供する仮想的な記憶領域 (即ち仮想化スイッチによって仮想化された記憶領域) を以下仮想ボリューム 100 と称する。

【0021】

仮想ボリューム 100 は、仮想化スイッチ 11 によって、少なくとも 1 つの実領域 132 と対応づけられる。仮想化スイッチ 11 は、複数の仮想ボリューム 100 をホストプロセッサ 12 等に提供することが出来る。仮想ボリューム 100 の各々には、仮想ボリュームを特定するための仮想化スイッチ 11 内で一意の識別子 (以下「仮想ボリューム識別子」) が与えられている。また、個々の仮想ボリューム 100 の記憶領域には連続したアドレス (以下仮想アドレスとも呼ぶ) が付されている。ホストプロセッサ 12 は、ストレージ装置 13 の LU 131 内の実領域 132 を直接指定する代わりに、仮想ボリューム識別子及び仮想ボリューム 100 内の場所を示す仮想アドレスを指定してアクセス要求を発行する。仮想化スイッチは、仮想ボリューム識別子及び仮想アドレスから、これに対応する LU 131 内の実領域 132 を特定し、特定された実領域 132 へのアクセス要求を発行する。従って、ホストプロセッサ 12 は、仮想ボリューム識別子と仮想ボリューム 100 内の記憶領域を示す仮想アドレスを用いて、ストレージ装置 13 の実領域 132 に格納されたデータを利用できる。

【0022】

管理コンソール 14 は、仮想ボリューム 100 を作成 (設定) するためにシステム管理者によって使用される計算機であり、表示装置と入力装置を備える。管理コンソール 14 は、仮想化スイッチ 11 とネットワークを介して接続されている。

【0023】

図 2 は、仮想化スイッチ 11 の内部構成例を示す図である。

【0024】

仮想化スイッチ 11 は、入力ポート 240、出力ポート 250、転送部 230、制御部 210 及び記憶部 220 を有する。

【0025】

入力ポート 240 は、仮想化スイッチ 11 がホストプロセッサ 12 と通信するための通信線と接続されるポートである。出力ポート 250 は、仮想化スイッチ 11 がストレージ装置 13 と通信するための通信線と接続されるポートである。尚、入力ポート 240 及び出力ポート 250 を構成する装置は同一であっても良い。この場合、どのポートを入力ポートあるいは出力ポートとして使用するかは、使用者が選択する。仮想スイッチ 11 は入力ポート 240 及び出力ポート 250 を各々若しくは複数有する。

【0026】

転送部230はメモリを有し、そのメモリに転送情報テーブル231を保持する。転送情報テーブル231には、入力ポート240と該入力ポート240を介して仮想化スイッチ11と通信可能なホストプロセッサ12との間の対応関係、及び出力ポート250と該出力ポート250を介して仮想化スイッチ11と通信可能なストレージ装置13との間の対応関係についての情報が格納される。

【0027】

転送部230は、転送情報テーブル231を参照し、入力ポート240がホストプロセッサ12から受信した入出力要求を、要求先のストレージ装置13と仮想化スイッチ11との間の通信に使用される出力ポート250へ転送する。また、転送部230は、出力ポート250がストレージ装置13から受信した応答情報やデータを、データ等を受信すべきホストプロセッサ12と仮想化スイッチ11の間の通信に使用される入力ポート240へ転送する。ただし、ホストプロセッサ12から受け取った入出力要求が仮想ボリューム100に対する入出力要求である場合には、後述のアクセス変換プログラム212を制御部210が実行した結果選択されるストレージ装置13へ、入出力要求を送信すべく、転送部230は、このストレージ装置13との間の通信に使用される出力ポート250に入出力要求を転送する。

【0028】

制御部210はプロセッサ及びメモリを有し、そのメモリに仮想ボリューム定義プログラム211、アクセス変換プログラム212及び割当処理プログラム213の各プログラムが格納される。これらのプログラムは、制御部210のプロセッサで実行される。

【0029】

記憶部220は、仮想ボリューム管理テーブル221、実領域管理テーブル222及びアクセス変換テーブル224を記憶する。

【0030】

アクセス変換テーブル224は、仮想化スイッチ11がホストプロセッサ12に提供する仮想ボリューム100ごとに存在する。アクセス変換テーブル224は、一又は複数のエントリ331と、対応する仮想ボリューム100の仮想ボリューム識別子を登録するエントリ332とを保持する。個々のエントリ331には、仮想ボリューム100内の記憶領域を示す仮想アドレス範囲と、その仮想アドレス範囲に対応する実領域132が属するLU131を指定する識別子であるLUアドレスと、当該実領域132のLU131内における位置を示すLU内アドレスとが対応付けられて登録される。即ちアクセス変換テーブル224は、仮想ボリューム100内の記憶領域を指定するための仮想アドレスと、ストレージ装置13内の記憶領域を指定するためのアドレスとの対応情報を保持している。

【0031】

なお、仮想ボリューム100内に実領域132が割当てられていない記憶領域が存在する場合、当該記憶領域に対応するエントリ331には、当該記憶領域の仮想アドレス範囲と、LUアドレスとして実領域132が割当てられていないことを示す「-1」、LU内アドレスとして同じく実領域132が割り当てられていないことを示す「-1」が登録される。アクセス変換テーブル224は、仮想ボリューム100の記憶領域の構成が変更されたとき、即ち仮想ボリューム100内の記憶領域に対応づけられている実領域132に変更があったとき、制御部210によって更新される。制御部210は、アクセス変換プログラム212を実行し、アクセス変換テーブル224を参照して、ホストプロセッサ12から受け取った仮想ボリューム100に対する入出力要求を、対応するストレージ装置13が有するLU131への入出力要求に変換する。また、制御部210は変換後の入出力要求のターゲットであるLU131を有する記憶装置13を転送部230に通知する。従って、転送部230は転送情報テーブル231を参照して、通知された記憶装置13と接続されている出力ポート250に、変換後の入出力要求を転送することになる。

【0032】

尚、ホストプロセッサから受信した入出力要求によって指定される仮想ボリューム100

0内の記憶領域に対応するLU131が存在しない場合（即ち、当該仮想ボリューム100内の記憶領域に対して未だ実領域が割り当てられておらず、アクセス変換テーブル224のエントリ331に、LUアドレス及びLU内アドレスとして「-1」が登録されている場合）、制御部210は、仮想ボリューム定義プログラム211を実行して、仮想ボリューム100の定義変更処理を実行する。

【0033】

性能向上のために、仮想化スイッチ11は、アクセス変換テーブル224およびアクセス変換プログラム212を各入力ポート240ごとに有していても良い。実領域管理テーブル222は、LU131ごとに存在する。実領域管理テーブル222は、LU131に含まれる実領域132を管理するために使用されるテーブルである。各実領域管理テーブル222には、ストレージ装置ID321、LUアドレス322及び実領域リスト324が格納される。

【0034】

ストレージ装置ID321は、LU131を保持するストレージ装置13を示す識別子である。LUアドレス322は、当該LU131の識別子である。

【0035】

実領域リスト324は、少なくとも1つのエントリ325を有する。各エントリ325は、LU131を構成する個々の実領域132に対応して設けられ、実領域ID、実領域のサイズ及び仮想ボリューム識別子の情報を登録する項目を有する。実領域IDはエントリ325に対応する実領域132を特定するための識別子、サイズは実領域132のサイズである。またエントリ325に登録される仮想ボリューム識別子は、実領域132が割り当てられている仮想ボリューム100を識別するための仮想ボリューム識別子である。実領域リスト324内のエントリ325は、実領域132のアドレス順に配列されている。

【0036】

なお、本実施形態では、実領域132のサイズは固定であると仮定するため、エントリ325に実領域132のサイズを登録する項目を含めなくても良い。また、未使用である実領域132については、エントリ325の仮想ボリューム識別子を登録する項目には、未使用であることを示すnullが登録される。

【0037】

このように、実領域管理テーブル222は、LU131に属する各実領域132が仮想ボリューム100として利用されているか否かを示す情報を保持しており、仮想化スイッチ11が仮想化ボリューム100に新たに割り当てる実領域132を選択する際に使用される。

【0038】

尚、実領域管理テーブル222は、仮想化スイッチ11にストレージ装置13が接続された時等のタイミングで、管理端末14を介した管理者の指示に基づいて作成される。さらに、この際LU131とそれを構成する実領域132の記憶容量等が決定される。尚、実領域管理テーブル222が作成された時点では、全てのエントリ325について、実領域IDには仮想化スイッチ11内で一意に決定される識別子が書き込まれ、仮想ボリューム識別子にはnullが設定される。

【0039】

仮想ボリューム管理テーブル221は、仮想ボリューム100ごとに存在する。各仮想ボリューム管理テーブル221には、識別子エントリ311及び実領域リスト315が格納される。識別子エントリ311には、仮想ボリューム管理テーブル221に対応する仮想ボリューム100の仮想ボリューム識別子が登録される。実領域リスト315は、仮想ボリューム管理テーブル221に対応する仮想ボリューム100にどの実領域132が割り当てられているかを示すリストである。実領域リスト315内では、仮想ボリューム100上の仮想アドレス順に対応する実領域132のエントリ317が並んでおり、各エントリ317には仮想アドレスに対応する実領域132の実領域IDが格納されている。当該仮想ボリューム100のうち、実領域132が割り当てられていない記憶領域に相当するエ

ントリ 3 1 7 には、有効な実領域 I D 3 1 7 の代わりに空を示す null 値が格納される。

【 0 0 4 0 】

このように、仮想ボリューム管理テーブル 2 2 1 は、仮想ボリューム 1 0 0 の記憶領域がどの実領域 1 3 2 と対応付けられているかを示す情報を保持しており、仮想化スイッチ 1 1 が解放可能な実領域 1 3 2 を選択する際に使用される。

【 0 0 4 1 】

以下、本実施形態における仮想化スイッチ 1 1 が行う記憶領域の割当て処理について説明する。

【 0 0 4 2 】

制御部 2 1 0 は、仮想ボリューム定義プログラム 2 1 1 を実行することで、仮想ボリューム 1 0 0 の定義を作成または変更する。制御部 2 1 0 は、管理コンソール 1 4 を介してシステム管理者から仮想ボリューム 1 0 0 の作成要求を受け取ると、仮想ボリューム定義プログラム 2 1 1 を実行して、仮想ボリューム管理テーブル 2 2 1 及びアクセス変換テーブル 2 2 4 を新規作成する。この場合、制御部 2 1 0 は、既に作成された他の仮想ボリューム 1 0 0 と重複しない仮想ボリューム識別子を生成して、これを仮想ボリューム管理テーブル 2 2 1 の識別子エントリ 3 1 1 に格納し、実領域リスト 3 1 5 内の各エントリ 3 1 7 に空の値 (null 値) を設定することで実領域リスト 3 1 5 を初期化する。このように、仮想ボリューム 1 0 0 の生成直後は、仮想ボリューム 1 0 0 に実領域 1 3 2 が対応付けられない。従って、このとき制御部 2 1 0 は、新規作成された仮想ボリューム管理テーブル 2 2 1 によって定義される仮想ボリューム 1 0 0 に対応するアクセス変換テーブル 2 2 4 についても、エントリ 3 3 2 には当該仮想ボリューム 1 0 0 の仮想ボリューム識別子を登録するが、エントリ 3 3 1 の LU アドレス及び LU 内アドレスには空の値 (「- 1」) を登録し、アクセス変換テーブル 2 2 4 を初期設定する。

【 0 0 4 3 】

このようにアクセスボリューム管理テーブル 2 2 1 とアクセス変換テーブル 2 2 4 が初期設定されていれば、ホストプロセッサ 1 2 が仮想化スイッチ 1 1 に接続されて仮想化スイッチ 1 1 に記憶領域を検出するためのコマンドを発行した際には、仮想化スイッチ 1 1 の制御部 2 1 0 は所定の大きさの仮想ボリューム 1 0 0 がホストプロセッサ 1 2 からアクセス可能であることを示す情報を通知する。この制御部 2 1 0 がホストプロセッサ 1 2 に通知する情報には、仮想ボリューム識別子と仮想ボリュームのサイズ (所定のサイズ) が含まれる。しかし、その仮想ボリューム 1 0 0 にはストレージ装置 1 3 が有する論理的な記憶領域である実領域 1 3 2 は実際には割当てられていない状態となる。仮想化スイッチ 1 1 は、その後、ホストプロセッサ 1 2 等からデータの書き込み要求を受信したタイミングで、仮想ボリューム 1 0 0 に実領域 1 3 2 を割当ててゐる。この結果、実際に計算機が記憶装置にアクセス要求を発行した際にこの計算機に対して実領域 1 3 2 を割り当てることができるので、予め容量に余裕を持って計算機に実領域 1 3 2 を割り当てておく必要がなく、計算機に割当てゐる記憶領域の無駄を省くことが出来る。

【 0 0 4 4 】

実領域 1 3 2 を仮想ボリューム 1 0 0 に割当てゐる際には、制御部 2 1 0 は、ホストプロセッサ 1 2 のデータ書き込み要求等の要求に従い、仮想ボリューム管理テーブル 2 2 1 を変更する。この場合、制御部 2 1 0 は、仮想ボリューム 1 0 0 に実領域 1 3 2 を割当てた後、アクセス変換テーブル 2 2 4 を更新する。尚、実際には、入力ポート 2 4 0 から受信した入出力要求の宛先が仮想ボリューム 1 0 0 であった場合に、転送部 2 3 0 は、その入出力要求を制御部 2 1 0 に転送する。制御部 2 1 0 は、転送された入出力要求について、図 3 に示す以下の処理を実行する。また、図 3 の処理終了後、制御部 2 1 0 は変換した入出力要求の要求先の情報を転送部 2 3 0 に送信し、転送部 2 3 0 はその情報に基づいて入出力要求を各ストレージ装置 1 3 に転送する。

【 0 0 4 5 】

図 3 は、制御部 2 1 0 がホストプロセッサ 1 2 から入出力要求を受け取った際に行う処理の一例を示すフローチャートである。

【0046】

まず、制御部210は、ホストプロセッサ12から受け取った仮想ボリューム100に対する入出力要求が書き込み要求であるかどうかを判断する（ステップ2001）。

【0047】

ホストプロセッサ12からの入出力要求が書き込み要求である場合、制御部210は、書き込み要求で指定された仮想ボリューム100の仮想アドレスに実領域132が対応付けられているかどうかを、アクセス変換テーブル224で確認する（ステップ2002）。

【0048】

指定された仮想ボリューム100の仮想アドレスに対応する実領域132がアクセス変換テーブル224に登録されていない場合（即ち入出力要求中に含まれる仮想ボリューム識別子と仮想アドレスに対応する、アクセス変換テーブル224のエントリ331にLUアドレス及びLU内アドレスとして「-1」が登録されている場合）、制御部210は、仮想ボリューム定義プログラム211を実行して仮想ボリューム100の定義変更処理を行う。

【0049】

具体的には、制御部210は、入出力要求中で指定された仮想ボリュームの仮想アドレス範囲に実領域132を割当てて、このために、制御部210はまず実領域管理テーブル222を検索して、未使用の実領域132（即ち、実領域管理テーブル222のエントリ325に仮想ボリューム識別子としてnull値が登録されている実領域132）を探し出す。そして制御部210は、探し出した未使用の実領域132を、入出力要求で指定された仮想ボリューム100の仮想アドレスに対応する記憶領域に割当てて、即ち制御部は、実領域管理テーブル222のうち、当該未使用の実領域132に対応するエントリ325に、入出力要求中に含まれる仮想ボリューム識別子を登録し、入出力要求中に含まれる仮想ボリューム識別子に対応する仮想ボリューム管理テーブル221のエントリ315に、探し出された未使用の実領域132の実領域IDを登録する。更に制御部は、アクセス変換テーブル224のうち、入出力要求中に含まれる仮想ボリューム識別子及び仮想アドレスに対応するエントリ331に、探し出された未使用の実領域132のLUアドレス及びLU内アドレスを登録する。

【0050】

尚、検索された未使用の実領域132の記憶容量が、入出力要求中で指定される仮想ボリューム100の仮想アドレス領域に対して不足する場合、制御部210は更に空きの（未使用の）実領域132を検索し、仮想アドレス領域に充当するまで検索を行う（ステップ2006）。

【0051】

ステップ2006の処理の後、若しくはステップ2002で入出力要求中で指定された仮想ボリュームの仮想アドレスに対応する実領域132が登録されていると判断された場合、若しくはステップ2001で書き込み要求ではないと判断された場合は、制御部210は、アクセス変換テーブル224を参照して、ホストプロセッサ12から受け取った仮想ボリューム100の仮想アドレスに対する入出力要求を、対応するストレージ装置13のLU131の実領域132への入出力要求に変換して（ステップ2009）、処理を完了する。

【0052】

尚、制御部210は、変換後の入出力要求を転送部230へ渡し、転送部230は変換後の入出力要求のあて先であるストレージ装置13に基づいて、転送情報テーブル231を参照し、この記憶装置13に変換後の入出力要求を、出力ポート250を介して送信する。

【0053】

このように図3のステップ2006の処理によって、仮想ボリューム100へデータが書き込まれる際に、データが書き込まれる仮想ボリューム上の記憶領域に対応する実領域

132を仮想ボリューム100に実際に割当てることができる。従って、計算機は初期設定当初から大きな容量を有する記憶領域（仮想ボリューム）が割り当てられていると認識しているが、実際には計算機が必要とする容量の実領域132を割り当てておけばよく、効率的な記憶領域の割当てが可能である。

【実施例2】

【0054】

次に、本発明の第二の実施形態を説明する。

【0055】

第一の実施形態では、仮想ボリュームに対して実領域132が割り当てられた後、その実領域132に格納されたデータが使用されなくなる場合についての考慮がされていない。そこで、第二の実施形態では、第一の実施形態に加え、使用されなくなった実領域132の仮想ボリューム100への割当てを止める（以下「解放する」と呼ぶ）ことを考慮する。

【0056】

本実施形態においては、第一の実施形態の構成に加え、制御部210のメモリには、デフラグ処理プログラム214が格納されている。制御部210は、システム管理者からのデフラグ処理開始指示を管理コンソール14を通して受け取り、デフラグ処理プログラム214を実行して、仮想ボリューム100に格納されたファイル等のデータを再配置する。

【0057】

具体的には、制御部210は、ストレージ装置13に格納されたファイルシステムの管理情報（ファイルシステムに格納されている各ファイルのデータ部が、仮想ボリューム100上のどこに配置されているかを示す仮想アドレスのリストの集合）を読み取って、読み出したファイルシステムの管理情報に基づいて、実領域132に格納されたデータがなるべく仮想ボリューム内の連続した記憶領域に格納されるよう、データを実領域132内の空きの領域にコピーし、管理情報に格納されている仮想アドレスをコピー元を示す値からコピー先を示す値へ書き換えることにより配置を変更する。なお、この再配置処理によるデータ破壊を防止するために、本実施形態ではシステム管理者は該仮想ボリューム100を使用するファイルシステムを、デフラグ処理開始指示の前にアンマウントしておく必要がある。

【0058】

その後、制御部210は、仮想ボリューム100内の記憶領域のうち、実領域132の割当てが不要になった記憶領域の仮想アドレス範囲を特定し、特定された仮想アドレス範囲に対応する実領域132を解放する（即ち、ファイルシステムの管理情報から、全てのファイルについて、そのデータ部が格納されている仮想ボリューム100上の記憶領域がわかるので、それらの記憶領域以外でかつ管理情報自身が格納されている仮想ボリューム100上の記憶領域以外の、仮想ボリューム100上の記憶領域に割当てられている実領域132を検索し、検索により特定された実領域132を解放する）。この際制御部210は、当該仮想ボリューム100についての仮想ボリューム管理テーブル221のうち、解放される実領域132の実領域IDが登録されているエン트리317にnull値を登録し、更に制御部210は、当該仮想ボリューム100に関するアクセス変換テーブル224のうち、解放される実領域132のLU131内での位置を示すLU内アドレスが登録されているエン트리331に、LUアドレス及びLU内アドレスとして「-1」を登録する。また制御部は、実領域管理テーブル222も更新し、解放される実領域132の実領域IDが登録されているエン트리325の仮想ボリューム識別子としてnull値を登録する。

【0059】

図4は、第二の実施形態において、仮想ボリュームへの割当てが不要となり解放可能となった実領域132を特定する処理の概念を示す図である。

【0060】

図4は仮想ボリューム100におけるファイルの配置及び仮想ボリューム100と実領

域 132 との対応関係を例示しており、図 4 (1) は制御部 210 のデフラグ処理プログラム 214 が実行される前の例を、図 4 (2) はデフラグ処理プログラム 214 が実行された後の例を示す。

【0061】

図 4 (1) において、仮想ボリューム 100 の図示された範囲の記憶領域には、3 個のファイル (ファイル A、ファイル B、ファイル C) が格納されており、ファイル A のデータは仮想ボリューム 100 内の記憶領域 501a1 及び記憶領域 501a2 に分割されて格納されており、ファイル B のデータは記憶領域 501b に、ファイル C のデータは記憶領域 501c に格納されている。一方、仮想ボリューム 100 の図示された範囲の記憶領域には、2 個の実領域 132 (実領域 132-1 と実領域 132-2) が対応付けられている。

【0062】

ここで制御部 210 はデフラグ処理プログラム 214 を実行して、ファイル A、ファイル B、ファイル C が、仮想ボリューム 100 内の連続する記憶領域にアドレス順に格納されるよう、これらのファイルを再配置する。

【0063】

その結果、再配置処理後は、図 4 (2) のように、仮想ボリューム 100 の記憶領域中左 (左の方がアドレスが小さいと仮定する) から順に、ファイル A のデータが記憶領域 501a1 及び記憶領域 501a2 に、ファイル B のデータが記憶領域 501b に、ファイル C のデータが記憶領域 501c に連続して配置される。すると、記憶領域 501c の後ろにはまとまった空き領域が出現する。この仮想ボリューム上での空き領域には実領域 132 を割当てておく必要がない。従って図 4 (2) の例では、仮想ボリューム上の空き領域に対応する実領域 132-2 は解放することが出来る。

【0064】

制御部 210 は、デフラグ処理プログラム 214 の実行後に出現した仮想ボリューム上の空き領域に対応する実領域 132 を、仮想ボリューム管理テーブル 221 から検索する。つまり制御部 210 は、仮想ボリューム管理テーブル 221 を検索して、空き領域の仮想アドレス範囲に対応する実領域を見つけ出す。そして見つかった実領域 132 を制御部 210 は解放する。具体的には、制御部 210 は解放する実領域 132 の実領域 ID を仮想ボリューム管理テーブル 221 から削除してこの実領域 ID が登録されていたエントリ 317 に null 値を登録し、実領域管理テーブル 222 内で解放される実領域 132 と対応付けられていた仮想ボリューム識別子を削除してこの仮想ボリューム識別子が登録されていたエントリ 325 に null 値を設定する。更に制御部 210 は、アクセス変換テーブル 224 に登録されている、解放される実領域 132 が属する LU 131 の識別情報 (LU アドレス) とこの LU131 内での解放される実領域 132 の位置情報 (LU 内アドレス) 情報を削除して代わりに「-1」を設定する。

【実施例 3】

【0065】

次に、本発明の第三の実施形態を説明する。

【0066】

第一の実施形態では、仮想ボリューム 100 に割当てられる実領域 132 のサイズは固定であった。しかし、所定の処理、例えばフォーマット処理に基づくデータの書き込み要求では、比較的記憶領域の使用量が小さい書き込み処理が発生する。この場合にも他の実施形態と同様に仮想ボリューム 100 に固定サイズの実領域 132 を割当てると、割当てた記憶領域に無駄が発生する。そこで、本実施形態では、処理内容に見合ったサイズの実領域 132 を仮想ボリューム 100 に割当てて、

【0067】

本実施形態は第一の実施形態の構成に加え、制御部 210 のメモリにフォーマット処理プログラム 215 が格納されている。また、仮想ボリューム管理テーブル 221 のエントリ 317 に、対応する実領域 132 のサイズを示す情報が含まれる。また、実領域管理テーブル 222 のエントリ 325 に含まれる実領域 132 のサイズは、本実施形態では第一

の実施形態と異なり省略されない。仮想ボリューム管理テーブル 221 のエントリ 317 及び実領域管理テーブル 222 のエントリ 325 にサイズを含めるのは、上述したように、実領域 132 のサイズを可変長として、割当効率を上げるためである。

【0068】

制御部 210 は、フォーマット処理プログラム 215 を実行することで、ホストプロセッサ 12 に代わり、仮想ボリューム 100 を使用するファイルシステムを初期化、即ちファイルシステム上のファイル及びディレクトリを全て消去し、新規にファイル及びディレクトリを作成可能な状態にする。このとき、制御部 210 は仮想ボリューム 100 に、メタデータと呼ぶ管理データを書き込むための入出力要求を自分自身に対して発行する。なお、制御部 210 は、この入出力要求の送信元アドレス（具体的には FibreChannel の Port ID）に制御部 210 自身を示す特別な値、具体的には 0xFFFF00 を含める。このとき書き込まれるメタデータのサイズはそれほど大きくないが、メタデータは仮想ボリューム 100 の記憶領域に一定の間隔に書き込まれるため、仮想ボリューム 100 のサイズが大きい場合は 1 つの仮想ボリュームに多数のメタデータが一定の記憶領域間隔で書き込まれる。従ってメタデータが仮想ボリュームに書き込まれるたびにメタデータサイズより大きな実領域 132 を当該仮想ボリュームに割り当てていると、仮想ボリュームに割り当てられた実領域 132 内にはデータが書き込まれていない未使用領域が発生し、実領域 132 の割当効率が低下する。実領域 132 のサイズを可変長とするのは、この割当効率の低下を防ぐためである。

【0069】

尚、フォーマット処理プログラム 215 は、管理者等の管理端末 14 を介した指示に基づき、あるいはホストプロセッサ 12 からの指示に基づき実行される。

【0070】

図 5 は、第三の実施形態において、ホストプロセッサ 12 から入出力要求を受信した場合、又は制御部 210 自身が入出力要求を生成した場合の、制御部 210 の処理の一例を示すフローチャートである。

【0071】

本処理手順には、第一の実施形態の図 3 で示した処理手順に加え、入出力要求がフォーマット処理に基づく入出力要求かどうかを判断する処理（ステップ 2003）、及びフォーマット処理に基づく入出力要求である場合に、仮想ボリューム 100 に割当てる実領域 132 のサイズを変更する処理（ステップ 2004）が含まれる。尚、その他のステップの処理は第一の実施形態と同様であるので、ここでは説明しない。

【0072】

制御部 210 は、処理対象の入出力要求が書き込み要求であって（ステップ 2001 で Y の場合）、入出力要求で指定された仮想ボリューム 100 の仮想アドレスに対応する実領域 132 がアクセス変換テーブル 224 に登録されていない場合（ステップ 2002 で N の場合）、その入出力要求がフォーマット処理に基づく入出力要求か否かを判別する。

【0073】

具体的には、制御部 210 は、入出力要求がフォーマット処理プログラム 215 の実行により制御部 210 自身によって生成されたものである場合（入出力要求の送信元アドレスが制御部 210 自身を示す値（0xFFFF00）である場合）はフォーマット処理に基づく入出力要求であると判断し、それ以外、即ちホストプロセッサ 12 から受信した入出力要求である場合はフォーマット処理に基づく入出力要求ではないと判断する（ステップ 2003）。

【0074】

処理対象の入出力要求が、フォーマット処理に基づく入出力要求であった場合には、制御部 210 は、仮想ボリューム 100 に割当てる実領域 132 のサイズを、他のデータ書き込み処理の際に仮想ボリューム 100 に対して割当てる実領域 132 のサイズの 1/3 と指定する。この除数はシステム管理者が任意に定めることができる。また制御部 210 は、フォーマット処理に基づく入出力要求があった際に仮想ボリューム 100 に割り当

てる実領域 132 のサイズを直接指定しても良い。

【0075】

さらに、制御部 210 は図 3 に示すステップ 2006 と同様に実領域管理テーブル 222 から未使用の実領域 132 を検索する。その際、指定された大きさの空き実領域 132 がない場合は、制御部 210 はそれより大きな実領域 132 を 2 個に分割する処理を、指定された大きさの未使用実領域 132 が生成されるまで繰り返す。制御部 210 は、指定されたサイズの未使用の実領域 132 が見つかったら、その実領域 132 を仮想ボリュームに割当て、図 3 のステップ 2006 で説明したのと同様の方法でアクセス変換テーブル 224、実領域管理テーブル 222 及び仮想ボリューム管理テーブル 221 を更新する。

【0076】

図 6 は、第三の実施形態における、フォーマット処理の概要を示す図である。

【0077】

図 6 は仮想ボリューム 100 のメタデータの配置及び実領域 132 との対応を示しており、図 6 (1) は実領域 132 のサイズが固定されていると仮定した場合、図 6 (2) は実領域 132 のサイズが可変である場合を示す。

【0078】

図 6 (1) 及び図 6 (2) では共に、仮想ボリューム 100 に対して 2 つのメタデータ 502 が書き込まれている。

【0079】

図 6 (1) では実領域 132 のサイズは固定されているので、フォーマット処理が実行された際、制御部 210 は、第一のメタデータを書き込むために実領域 132-1 を、第二のメタデータを書き込むために実領域 132-2 を、各々仮想ボリュームに割り当てる必要がある。図 6 (1) に示すように、メタデータ 502 のサイズが、実領域 132 のサイズより小さい場合には、仮想ボリューム 100 に割り当てられた実領域内に、実際にはデータが書き込まれない空き領域が発生し、記憶領域の効率的な使用の妨げとなる。

【0080】

一方図 6 (2) では実領域 132 のサイズは 1 つには限られず可変であるので、フォーマット処理が実行された際、第一のメタデータを書き込むためには図 6 (1) の実領域よりサイズの小さい実領域 132-1 を、第二のメタデータを書き込むためには同じくサイズの小さい実領域 132-2 が仮想ボリューム 100 に割り当てられる。そして、仮想ボリューム 100 の他の部分については、実際にホストプロセッサからデータの書き込み要求が発行された際に、実領域 132-3 や実領域 132-4 を割り当てれば良い。従って、図 6 (1) の場合と比べて、仮想ボリュームに割り当てられた実領域内に、空き領域が発生し難い。

【0081】

このように、図 5 のステップ 2004 で仮想ボリューム 100 に割り当てる実領域 132 のサイズを縮小することで、仮想ボリューム 100 に割当てた実領域 132 中に実際には使用されない空き記憶領域が発生する確率を小さくし、記憶領域を効率的に割り当てることができる。

【実施例 4】

【0082】

次に、本発明の第四の実施形態を説明する。

【0083】

本実施形態では、第二の実施形態と同様に、制御部 210 は、一度仮想ボリューム 100 に割り当てられた実領域 132 のうち、解放可能な実領域 132 を検索し解放する。ただし、実領域の解放のために本実施形態では、制御部 210 はジャーナルファイルシステムにおけるログ情報を使用する。

【0084】

本実施形態では、ホストプロセッサ 12 でジャーナルファイルシステムが使用される。また本実施形態では、第一の実施形態の構成に加え、制御部 210 のメモリにジャーナル

解析プログラム 216 が格納され、仮想ボリューム管理テーブル 221 に、ジャーナル領域の仮想アドレス範囲を登録するエントリ 318 及びメタデータの複製を登録するエントリ 319 が含まれる。

【0085】

本実施形態においては、制御部 210 はジャーナル解析プログラム 216 を実行して、仮想ボリューム 100 に配置されたジャーナル領域に書き込まれた情報を解析し、解放可能な実領域 132 がある場合に実領域 132 の解放処理を行う。

【0086】

仮想ボリューム管理テーブル 221 内の、ジャーナル領域の仮想アドレス範囲が登録されるエントリ 318 は、入出力要求がジャーナル領域への書き込み要求か否かを判定するために制御部 210 によって参照される。仮想ボリューム 100 がジャーナルファイルシステムとしてフォーマットされていない場合、仮想ボリューム管理テーブル 221 内のジャーナル領域の仮想アドレス範囲エントリ 318 は空である（即ち当該エントリ 318 には null 値が登録されている）。仮想ボリューム 100 がジャーナルファイルシステムとして初期化されるときには、管理者が管理コンソール 14 を介して、ジャーナルファイルシステムのジャーナルが格納される仮想ボリューム中の仮想アドレス範囲を、仮想ボリューム管理テーブル 221 のエントリ 318 に書き込む。尚、ジャーナルが格納される仮想ボリューム中の仮想アドレス範囲は、ジャーナルファイルシステムのフォーマットプログラム（フォーマットプログラムは、制御部 210 のメモリに格納されている）が制御部 210 に実行されることによって、仮想ボリューム管理テーブル 211 のエントリ 318 に書き込まれることとしても良い。

【0087】

仮想ボリューム管理テーブル 211 内の、メタデータの複製エントリ 319 は、制御部 210 が仮想ボリューム 100 に格納されるメタデータの複製を保存する際に使用される。

【0088】

図 7 は、第四の実施形態において、入出力要求を受け付けた制御部 210 が実行する処理の一例を示すフローチャートである。

【0089】

本処理では、第一の実施形態の図 3 に示した処理に加え、入出力要求がジャーナル領域への書き込みか否かを判断する処理（ステップ 2007）、及び解放可能な実領域 132 の解放を行う処理（ステップ 2008）が、制御部 210 によって行われる。

【0090】

ステップ 2007 において制御部 210 は、処理対象の書き込み要求が、ジャーナル領域への書き込み要求か否かを判断する。この際制御部 210 は、書き込み要求中に含まれる書き込み先のアドレスが、仮想ボリューム管理テーブル 221 のエントリ 318 に登録されている仮想アドレス範囲に含まれるか否かに基づいて、書き込み要求がジャーナル領域への書き込み要求か否かを判断する。書き込み先のアドレスが仮想ボリューム管理テーブル 221 のエントリ 318 に登録された仮想アドレス範囲に入る場合には、書き込み要求はジャーナル領域に対する書き込み要求であるため、ジャーナル解析プログラム 216 の処理、即ちステップ 2008 へ進む。

【0091】

ステップ 2008 において制御部 210 は、ジャーナル解析プログラム 216 を実行してジャーナル領域に書き込まれる情報を解析する。ジャーナル領域に書き込まれる情報とは具体的には、メタデータの一部と当該一部のメタデータのメタデータ全体における（メタデータ内での）オフセットである。そして、ジャーナル領域に書き込まれる情報を元に制御部 210 はメタデータを構築して、仮想ボリューム管理テーブル 221 のエントリ 319 に格納する。

【0092】

メタデータにはファイルシステム上のファイルと、ファイルが格納されている仮想ボリ

ューム上の記憶領域の仮想アドレスとの対応情報が含まれている。従って制御部 2 1 0 はメタデータを参照することによって、ファイルが格納されていない仮想ボリューム内の記憶領域の仮想アドレスを特定することができる。そして制御部 2 1 0 は、ファイルが格納されていない（即ち、実領域 1 3 2 を割り当てる必要のない）仮想ボリューム上の記憶領域の仮想アドレス範囲を特定すると、その特定された仮想アドレス範囲を元に仮想ボリューム管理テーブル 2 2 1 を検索し、当該仮想アドレス範囲に対応する実領域 1 3 2 を特定する。そして、特定された実領域 1 3 2 を解放すべく、仮想ボリューム管理テーブル 2 2 1、アクセス変換テーブル 2 2 4 及び実領域管理テーブル 2 2 2 を更新する。尚、仮想ボリューム管理テーブル 2 1 1、アクセス変換テーブル 2 2 4、実領域管理テーブル 2 2 2 の更新処理は第二の実施形態と同様であるので説明を省略する。

【0 0 9 3】

図 8 は、第四の実施形態における、実領域 1 3 2 の解放処理の概要を示す図である。

【0 0 9 4】

図 8 は、仮想ボリューム 1 0 0 におけるジャーナル領域、メタデータ、ファイルの配置及び仮想ボリューム 1 0 0 と実領域 1 3 2 との対応を示している。

【0 0 9 5】

図 8（１）は、ジャーナルファイルシステム上のファイル 5 0 1 のサイズが縮小される前を、図 8（２）はファイル 5 0 1 が更新されてファイルサイズが縮小された後の様子を示している。

【0 0 9 6】

ファイル 5 0 1 のサイズが更新によって縮小される場合、縮小されるファイル 5 0 1 を管理しているメタデータ 5 0 2 は、ファイルの更新時にホストプロセッサ 1 2 によって更新される。それに先立ち、ホストプロセッサ 1 2 は処理内容をジャーナルログとして記録するために、仮想ボリューム 1 0 0 内のジャーナル領域 5 0 3 に情報を書き込む。このジャーナル領域 5 0 3 には、メタデータ 5 0 2 の内容も書き込まれるので、ジャーナル領域 5 0 3 に書き込まれたメタデータ 5 0 2 を読み出せば、制御部 2 1 0 は読み出したメタデータの内容からファイル 5 0 1 が更新されてサイズが縮小され、更新後のファイルは図 8（２）に示した記憶領域に格納されることが分かる。

【0 0 9 7】

その結果、図 8（２）の場合は、実領域 1 3 2 - 4 に対応する仮想ボリューム 1 0 0 上の記憶領域が空きになることが分かるので、制御部 2 1 0 は仮想ボリューム管理テーブル 2 1 1 を用いてこの記憶領域に対応する実領域 1 3 2 - 4 が解放可能である旨を特定できる。

【0 0 9 8】

なお、メタデータ 5 0 2 にはファイルシステム内の全てのファイル 5 0 1 各々について、仮想ボリューム 1 0 0 内のどの記憶領域に格納されているかを示す情報が登録されている。従って、制御部 2 1 0 は、メタデータ 5 0 2 に登録された情報から、仮想ボリューム 1 0 0 内の記憶領域の内空き状態にある記憶領域を判断できる。

【0 0 9 9】

さらに制御部 2 1 0 は、ジャーナル領域 5 0 3 へデータが書き込まれると、ジャーナル領域 5 0 3 に書き込まれたデータから構築した最新のメタデータを仮想ボリューム管理テーブル 2 2 1 内のエントリ 3 1 9 に保存し、エントリ 3 1 9 に保存された最新のメタデータを参照して実領域 1 3 2 - 4 を解放する。

【0 1 0 0】

以上に述べた、（１）ファイル 5 0 1 の更新によるのファイルサイズの縮小、（２）ファイルの更新に伴うジャーナル領域 5 0 3 への情報の書き込み、（３）ジャーナル領域 5 0 3 に書き込まれたデータに基づく最新メタデータの生成及び、生成された最新メタデータによる仮想ボリューム管理テーブル 2 1 1 のエントリ 3 1 9 の更新、（４）実領域 1 3 2 - 4 の解放、はこの順で行われることになる。

【0 1 0 1】

尚、ホストプロセッサ12によるメタデータ502の書き込みは、ジャーナル領域503への情報の書き込み以降に、制御部210の処理とは非同期に行われる。そのために本実施形態では、制御部210はホストプロセッサによって仮想ボリューム100に書き込まれるメタデータ502を読み込むのではなく、代わりにジャーナル領域503に書き込まれたデータから生成され、仮想ボリューム管理テーブル211のエントリ319に格納されたメタデータを使用する。

【0102】

尚、上述の第四の実施形態においては、ファイルの更新によりファイルのサイズが縮小される場合を例に説明をしたが、ファイル501の削除によって実領域132が解放可能になる場合にも、同様の方法によって実領域132の解放が可能である。ファイル501が削除される場合にも、ジャーナル領域にはその削除処理についての内容（メタデータ）が記録されるので、制御部210は上述した処理と同様の方法で実領域132を解放することができる。

【0103】

このように、制御部210は、仮想ボリューム100内の記憶領域で、実際にはデータが格納されていない記憶領域に対応する実領域132を解放できる。

【実施例5】

【0104】

次に、本発明の第五の実施形態を説明する。

【0105】

本実施形態では、上述の第二、第三の実施形態において仮想化スイッチ11で実行されていたデフラグ処理プログラム214やフォーマット処理プログラム215等が専用の計算機にて実行される。

【0106】

本実施形態では、図1で示した第一の実施形態の構成に、仮想化スイッチ11に接続される計算機である専用サーバ15が追加される。専用サーバ15は、制御部210に代わり、デフラグ処理プログラム214およびフォーマット処理プログラム215の実行を行う。制御部210の処理能力やメモリ容量には限りがあり、仮想ボリューム100の数や対応するファイルシステムの種類には限界がある。そのため、専用サーバ15でデフラグ処理やフォーマット処理を代行する。

【0107】

したがって、第五の実施形態においては、仮想化スイッチ11の制御部210のメモリには、上述した実施形態で説明したデフラグ処理プログラム214やフォーマット処理プログラム215は格納されておらず、その代わり、仮想化スイッチ11に接続される専用サーバ15と仮想化スイッチ11との間の遣り取りを制御部210が制御するために実行されるプログラムであるサーバ連携プログラム217がメモリに格納されている。

【0108】

制御部210は、デフラグ処理を行うときに、デフラグ処理プログラム214を実行する代わりに、サーバ連携プログラム217を実行し、専用サーバ15にデフラグ処理開始要求を送信し、専用サーバ15から仮想ボリューム100上で空きになった記憶領域のリストを受信する。次に制御部210は、この空きになった記憶領域のリストから対応する実領域132を仮想ボリューム管理テーブル221を検索して特定し、特定された実領域132を解放する。

【0109】

制御部210は、フォーマット処理を行うときに、フォーマット処理プログラム215を実行する代わりにサーバ連携プログラム217を実行し、専用サーバ15にフォーマット処理開始要求を送信する。その後制御部210の処理は第三の実施形態と同様である。

【0110】

このように、本実施形態では専用サーバ15でデフラグ処理やフォーマット処理等を実行することにより、仮想化スイッチ11の処理負荷を軽くすることができるので、仮想化

スイッチ 11 が管理する仮想ボリューム 100 の数や、仮想化スイッチ 11 が対応可能なファイルシステムの種類を増やすことが出来る。

【実施例 6】

【0111】

図 9 から図 10 を用いて、本発明の第六の実施形態を説明する。

【0112】

本実施形態では、第一の実施形態に対して、仮想化スイッチ 11 がホストプロセッサ 12 に対して未使用の実領域 132 の合計サイズを報告することが異なる。これは以下の理由による。

【0113】

本発明ではホストプロセッサ 12 に対し、仮想ボリューム 100 の作成時または利用開始時に、所定の大きさの記憶領域（仮想ボリューム 100）がホストプロセッサ 12 に割り当てられ利用可能となったことを示す情報が通知されるが、実際には当該ホストプロセッサに割り当てられた仮想ボリューム 100 全体には実領域 132 は割り当てられていない。従って、後にホストプロセッサからの書き込み要求の受領時等に、新たにその仮想ボリューム 100 に実領域 132 を割り当てようとしても、未使用の実領域 132 が不足して十分なサイズの実領域 132 を仮想ボリューム 100 に割り当てられない可能性がある。

【0114】

そこで本実施形態では、割り当て可能な空きの（未使用の）実領域 132 のサイズの合計を管理者が確認可能とすることにより、ストレージ装置 13 を追加するなどして、仮想ボリューム 100 への実領域 132 の割り当て時に空きの実領域 132 が不足する事態を回避できるようにする。

【0115】

図 9 は、仮想化スイッチ 11 が記憶部 220 内に有する、空き容量管理テーブル 225 の一例を示している。

【0116】

空き容量管理テーブル 225 は仮想ボリューム識別子 332 と空き容量 226 とを列挙したテーブルである。空き容量 226 は、対になっている仮想ボリューム識別子 332 で特定される仮想ボリューム 100 に対して割り当て可能な空きの（即ち未使用の）実領域 132 のサイズの合計である。空き容量 226 は、システム管理者が個別に設定する。尚、実領域管理テーブル 222 に登録されている全ての空きの実領域 132 のサイズの合計を仮想ボリューム 100 の総数で等分した値を、空き容量管理テーブル 225 の空き容量 226 に設定しても、またはシステム管理者が設定した空き実領域 132 の各仮想ボリューム 100 への配分比に応じて空き容量 226 を設定しても良く、空き容量 226 の合計が実領域管理テーブル 222 に登録されている全ての空きの実領域 132 のサイズの合計を超えなければ、各仮想ボリューム 100 の空き容量 226 の設定方法は任意でよい。

【0117】

この空き容量管理テーブル 225 はホストプロセッサ 12 から、ある特別な仮想ボリューム（以下、空き容量管理仮想ボリュームと呼ぶ）として参照可能とし、その仮想ボリューム識別子として、仮想ボリューム 100 に割り当てられている仮想ボリューム識別子とは別の仮想ボリューム識別子（以下、空き容量管理仮想ボリューム識別子と呼ぶ）を付与しておく。ホストプロセッサは、通常の入出力要求と同様の形式のコマンドを用いて、空き容量管理仮想ボリューム識別子を指定して空き容量管理テーブル 225 の読み出し要求を仮想化スイッチに送信する。これを受信した仮想化スイッチは、空き容量管理テーブル 225 の内容を返す。

【0118】

なお、空き容量管理仮想ボリューム識別子は後述する空き容量表示プログラムに直接書き込まれるか、空き容量表示プログラムが参照するホストプロセッサ上の設定ファイルにシステム管理者が書き込む。

【0119】

また、仮想化スイッチ11とホストプロセッサ12がLANなど他の通信手段で互いに通信可能であれば、その通信手段に従った方法によってホストプロセッサ12は空き容量管理テーブル255の読み出しを行っても良い。

【0120】

図10は、制御部210がホストプロセッサ12から入出力要求を受取った際に行う処理の手順を示すフローチャートである。

【0121】

図3と異なり、図10において制御部210は、ステップ2006を実行する代わりにステップ2010を実行する。ステップ2010において制御部210は、仮想ボリューム定義プログラム211を実行して図3のステップ2006と同様に仮想ボリューム100の定義変更処理を実行すると共に更に空き容量管理テーブル225の更新処理を行う。

【0122】

具体的には、ステップ2006の処理に加えて制御部210は、空き容量管理テーブル225内を検索して、書き込み要求の対象となっている仮想ボリューム100の仮想ボリューム識別子332を探す。そして制御部210は、この仮想ボリューム識別子332に対応する空き容量226の値から、ステップ2010においてこの仮想ボリュームに対して割当てた実領域132の合計サイズを減算し、その値を空き容量226に書き戻す。

【0123】

一方、ホストプロセッサ12は内蔵するメモリに空き容量表示プログラムを格納している。ホストプロセッサ12に内蔵されたCPUは、空き容量表示プログラムを実行して、管理者に対し、ホストプロセッサ12が使用している仮想ボリューム100に今後割当て可能な実領域の容量を表示する。この「今後割当て可能な実領域の容量」として、仮想化スイッチからホストプロセッサが受信した、空き容量管理テーブル225の情報の内、当該仮想ボリューム100の空き容量226の値を使用が使用される。

【0124】

以上に述べた通り本実施形態では、管理者が仮想ボリューム100に割り当て可能な実領域132のサイズの合計を知ることができる。

なお、本実施例に示した方法は、上述した第二から第五の実施形態に対しても適用することができる。尚、上述の実施例1から実施例6に示した実施形態では、仮想化を実現する装置としてスイッチを例にあげて説明したが、それ以外の装置、例えば計算機でもルーターでも良い。

【図面の簡単な説明】

【0125】

【図1】 本発明を適用したシステムの全体構成の一例を示す図である。

【図2】 仮想化スイッチ11の内部構成の一例を示す図である。

【図3】 制御部210の入出力要求の処理手順の一例を示すフローチャートである。

【図4】 第二の実施形態におけるデータの再配置の処理概要の一例を示す図である。

【図5】 第三の実施形態における制御部210の入出力要求の処理手順の一例を示すフローチャートである。

【図6】 第三の実施形態におけるフォーマット処理の一例を示す図である。

【図7】 第四の実施形態における制御部210の入出力要求の処理手順の一例を示すフローチャートである。

【図8】 第四の実施形態における、実領域132の解放処理の一例を示す図である。

【図9】 第六の実施形態における、空き容量管理テーブル225の一例を示す図である。

【図10】 第六の実施形態における、制御部210の入出力要求の処理手順の一例を示すフローチャートである。

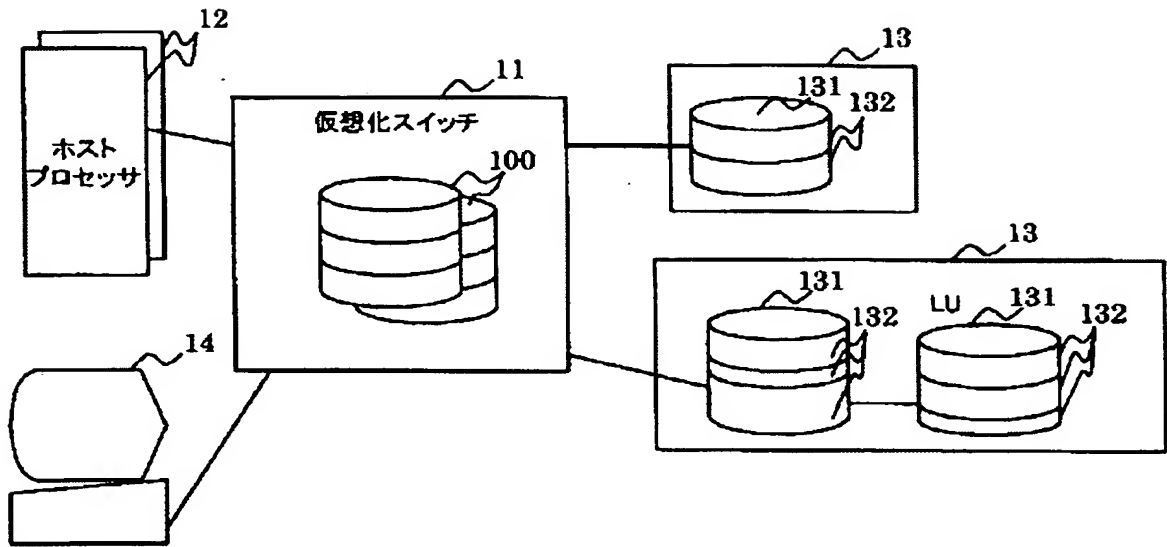
【符号の説明】

【0126】

- 1 1 …仮想化スイッチ
- 1 2 …ホスト
- 1 3 …ストレージ装置
- 2 1 0 …制御部
- 2 2 0 …記憶部

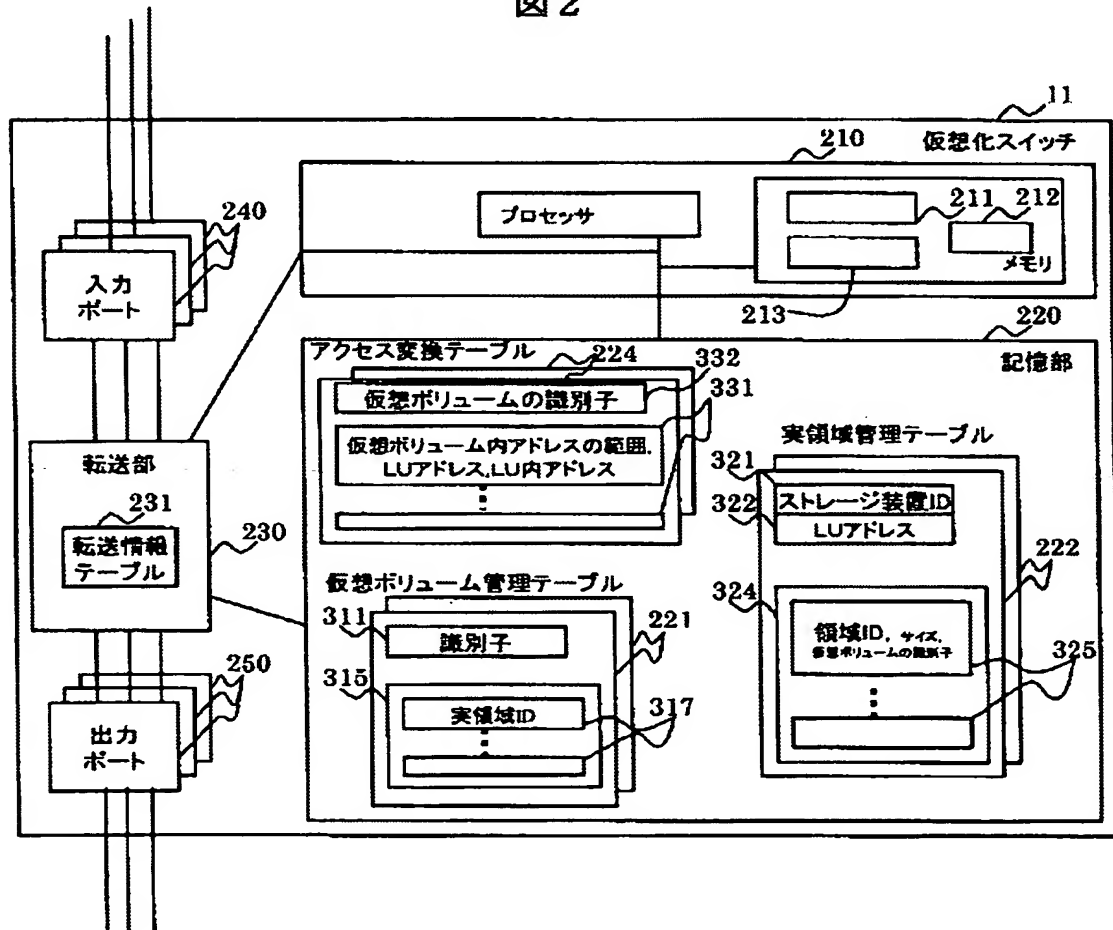
【書類名】 図面
【図 1】

図 1



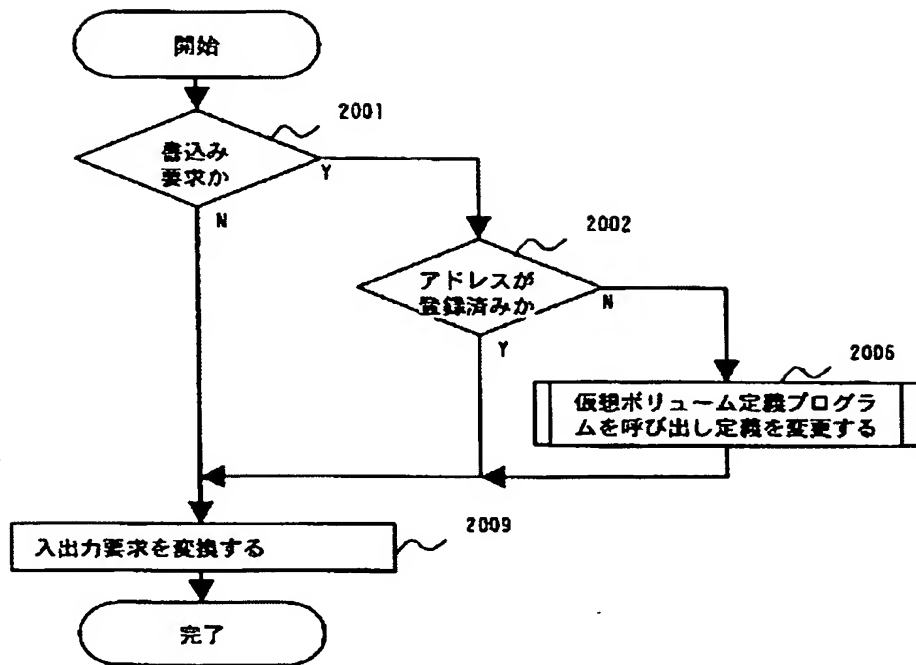
【図 2】

図 2



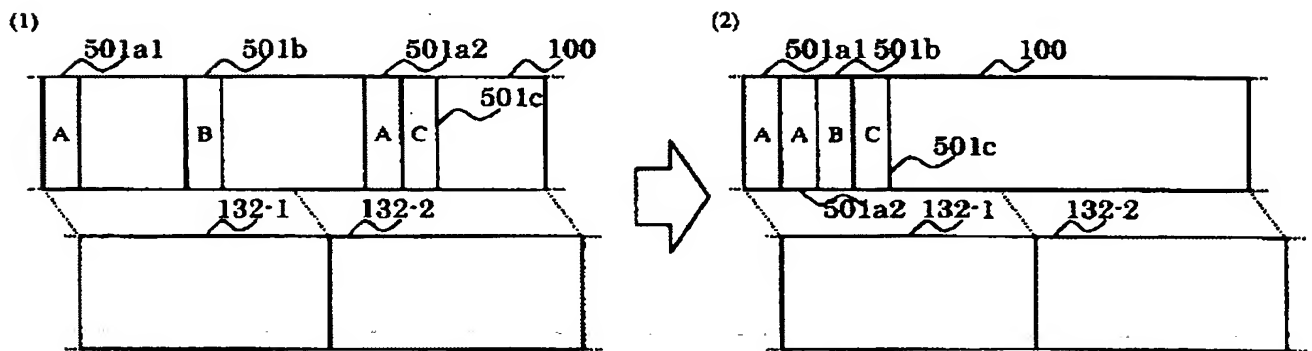
【図 3】

図 3



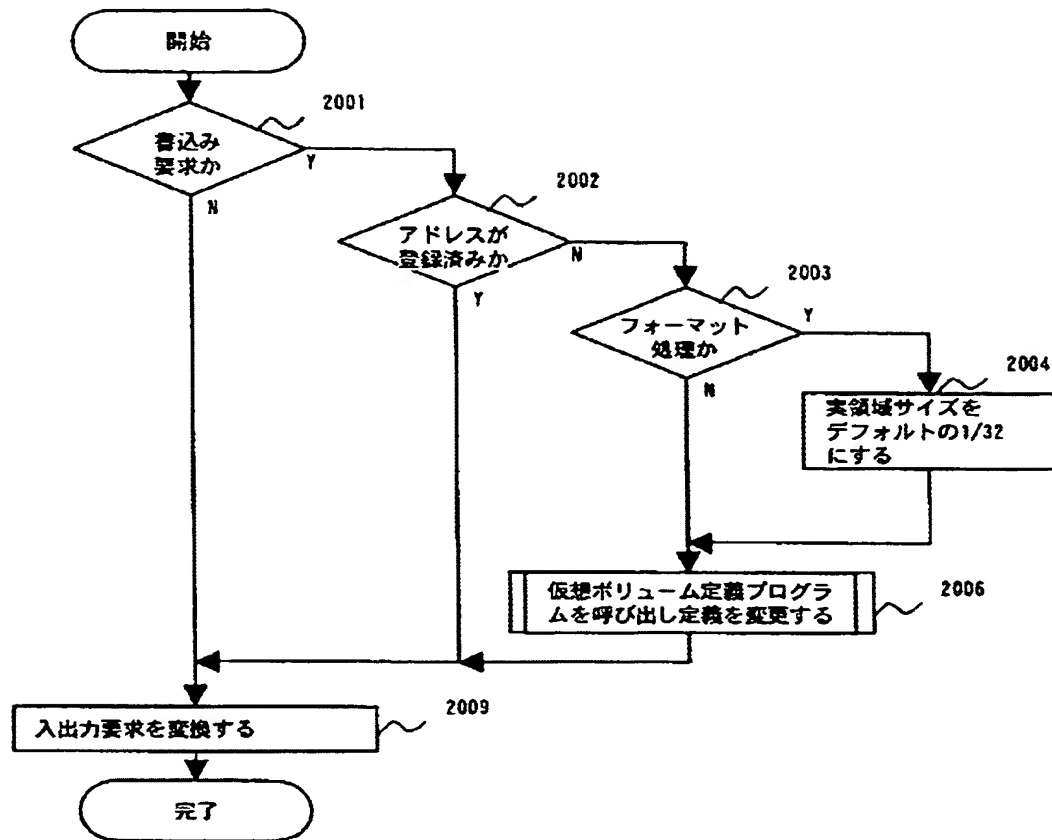
【図 4】

図 4



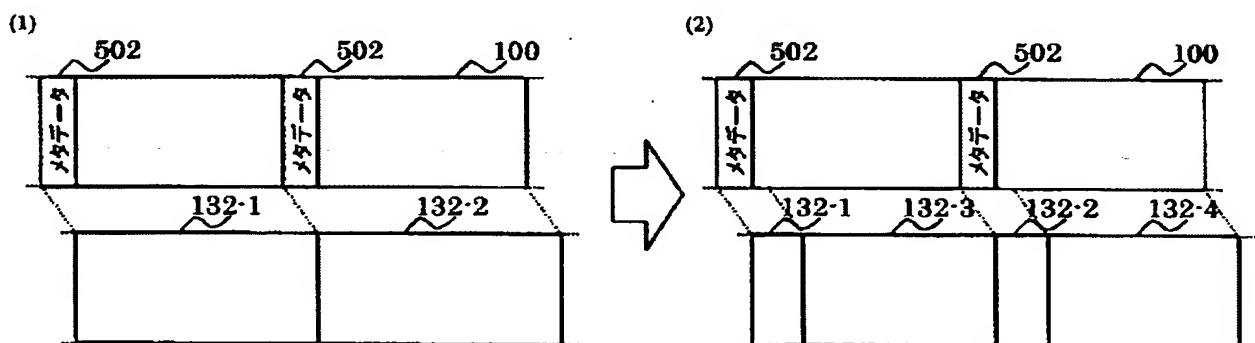
【図 5】

図 5



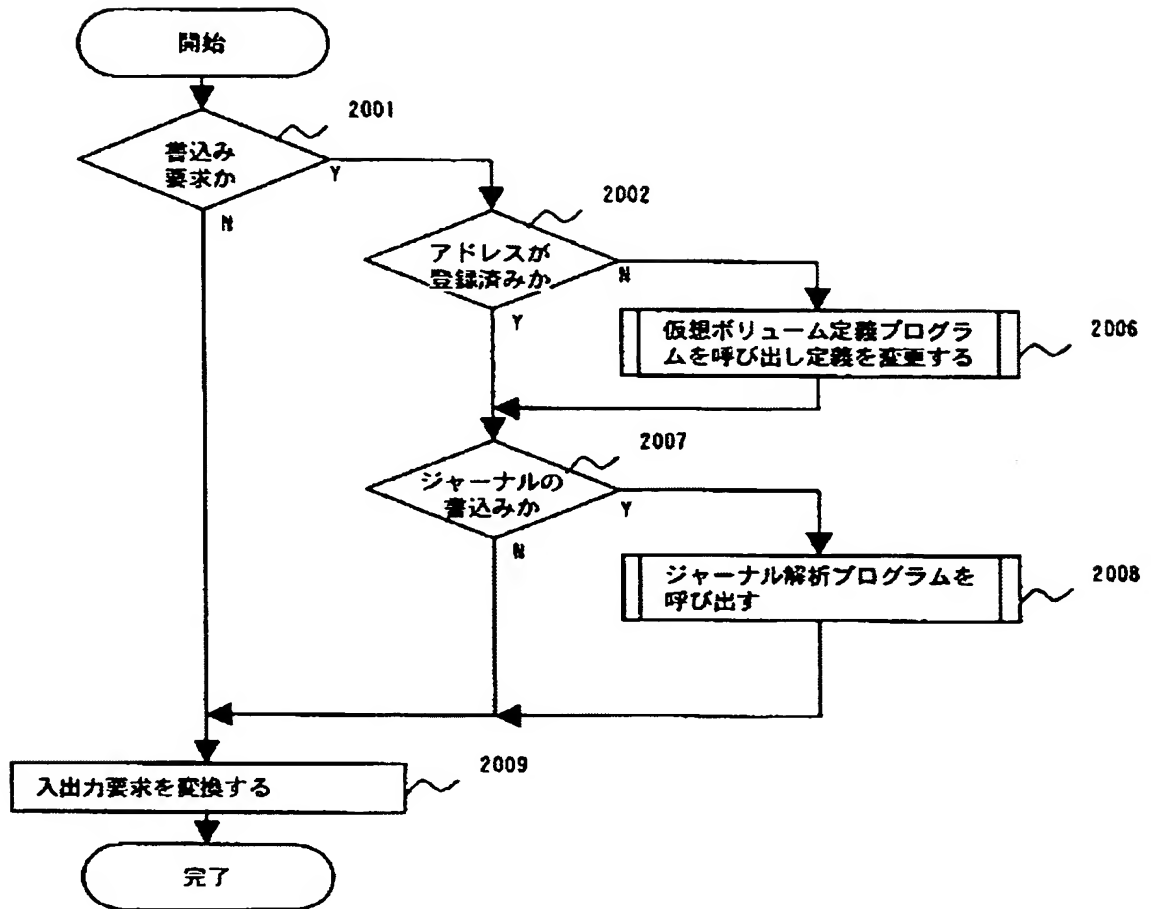
【図 6】

図 6



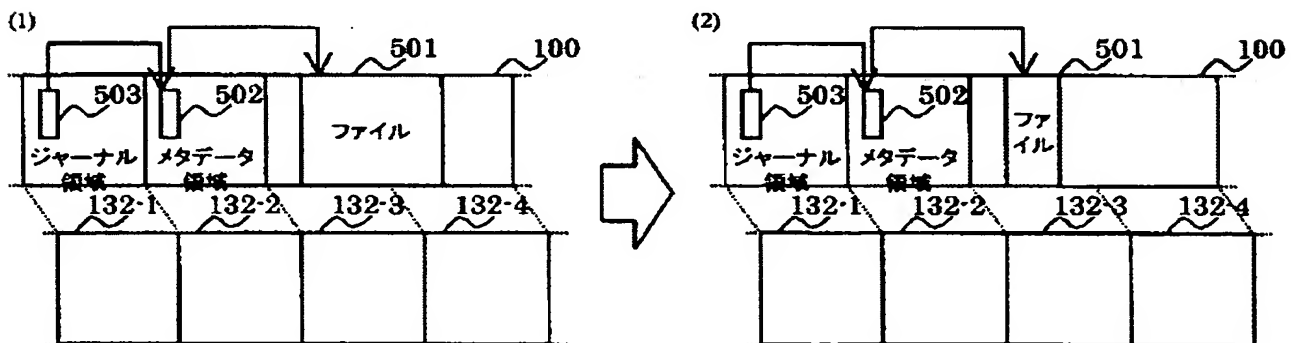
【図 7】

図 7



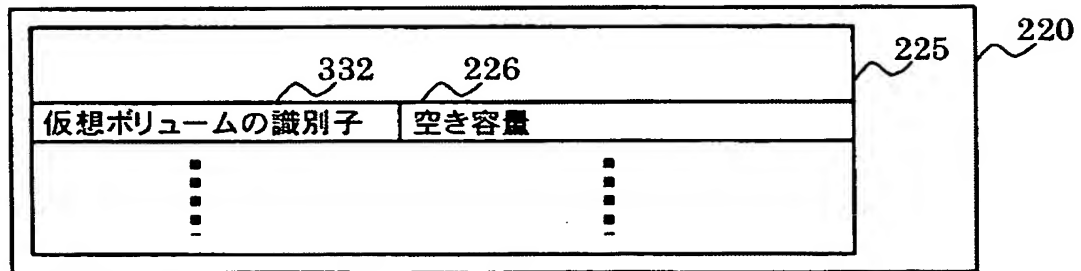
【図 8】

図 8



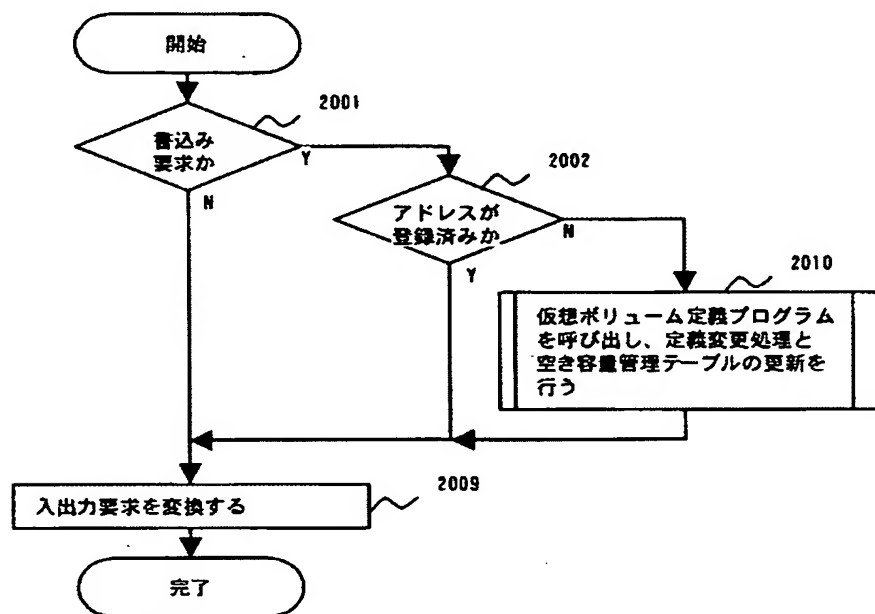
【図 9】

図 9



【図 10】

図 10



【書類名】 要約書

【要約】

【課題】

計算機に記憶装置が接続されたシステムにおいて、記憶装置内の記憶領域を計算機に効率的に割り当てる。

【解決手段】

仮想化スイッチ 1 1 は、仮想ボリュームの作成時に、仮想ボリュームに実領域を割り当てず、後にホストプロセッサ 1 2 が仮想ボリュームへ書き込みを行ったときに初めて、書き込みが行われた仮想ボリューム上の領域に実領域を割り当てる。

さらに、仮想化スイッチ 1 1 は、割当が不要となった実領域を解放する。また仮想化スイッチ 1 1 は、入出力要求の内容に応じて、割り当てる実領域の割当単位を小さくする。

【選択図】 図 2

認定・付加情報

特許出願の番号	特願 2 0 0 4 - 0 6 3 3 9 2
受付番号	5 0 4 0 0 3 7 3 3 5 2
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 6 年 3 月 1 1 日

< 認定情報・付加情報 >

【提出日】 平成16年 3月 8日

特願 2 0 0 4 - 0 6 3 3 9 2

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 5 1 0 8]

1. 変更年月日 1 9 9 0 年 8 月 3 1 日
[変更理由] 新規登録
住 所 東京都千代田区神田駿河台 4 丁目 6 番地
氏 名 株式会社日立製作所
2. 変更年月日 2 0 0 4 年 9 月 8 日
[変更理由] 住所変更
住 所 東京都千代田区丸の内一丁目 6 番 6 号
氏 名 株式会社日立製作所